December 20, 2024

E-Filed

The Honorable Thomas S. Hixson
United States District Court for the Northern District of California
San Francisco Courthouse, Courtroom E – 15th Floor
450 Golden Gate Avenue
San Francisco, CA 94102

Re: *Kadrey et al. v. Meta Platforms, Inc.*; Case No. 3:23-cv-03417-VC-TSH

Dear Judge Hixson:

    The parties jointly submit this letter brief regarding issues related to Plaintiffs' motion to compel as to Meta's Objections and Responses to Plaintiffs' First and Second Sets of Requests for Admission ("RFAs"), initially served on November 8 and November 18, and then amended in part on December 13, 2024. The parties met and conferred on December 17 but were unable to reach a resolution.

I.   PLAINTIFFS' POSITION

"The dual purpose of [RFAs] is to facilitate proof with respect to issues that cannot be eliminated from the case and to narrow the issues by eliminating those that can be." *I-Enterprise Co. v. Draper Fisher Jurvetson Mgmt. Co.*, 2005 WL 8177424, at *1 (N.D. Cal. Mar. 23, 2005) (cleaned up). Despite a prior Court order requiring Meta to amend many of its evasive RFA responses, Meta has once again failed to answer Plaintiffs' RFAs as written.

A. Meta Fails To Respond To RFAs 3-7, 17, 20, 23, 34, 43, 45-91, 94, and 96 As Written—They Ask About Books, Not Some Unidentified Amount of "Text."

This Court already held that Plaintiffs' "RFAs 3-7, 17, 20, 23, 34, and 43 ask about 'copyrighted books,' 'copyrighted works,' and 'copyrighted material," and that RFAs 45-91, 94, and 96 ask about "particular works." Dkt. 315 at 4, 6. And the Court admonished Meta for being "evasive and [] not answer[ing] these RFAs as required." *Id.* at 4. Yet Meta continues to be evasive—rather than answer the RFAs directly, Meta persists in rewriting and admitting them only as to unidentified "text" in copyrighted books, works and material, and "some text" in particular works. *See* Ex. A. The Court's guess is as good as Plaintiffs' as to whether "text" is a sentence, paragraph, chapter, or entirety of the copyrighted books, works, and material the RFAs ask about. The reason Meta does so is obvious—referencing unidentified "text" fits better with its fair use argument that Meta only cares about "text" generally and not any particular copyrighted work. But Meta does not get to rewrite Plaintiffs' requests to try to fit its theory of the case.

Further, Meta already told this Court that it "reasonably investigated whether Plaintiffs' at-issue works"—not "some text"—"are included in the Books3 dataset using published, commercially available copies of Plaintiffs' works," Dkt. 293 at 3, and that Meta "underwent significant burden in comparing the texts from the published, commercially-available versions of Plaintiffs' works against third-party datasets to determine whether the datasets included text from Plaintiffs' at-issue books (albeit not the deposit copies, which, as noted, Plaintiffs have failed to provide)." *Id.* at 4. Thus, Meta already knows and can admit the RFAs as written.

B. Meta Should Answer RFAs 7, 16, 19, 22, 26, 35, and 39 As Written.

This Court already held that <u>RFA 7</u> ("Admit that You did not obtain permission or consent from the relevant copyright owners to use all copyrighted books in the Datasets used to train Llama Models.") is "perfectly clear," and defined the "relevant copyright owners" as "the ones who own the copyrights to the books in the datasets used to train Llama models." Dkt. 315 at 6. Despite the Court's order, Meta redefined "relevant copyright owner" to mean "the person identified in the book as the copyright owner, without admitting that such person in fact owns a valid copyright in the book or what it covers." Ex. A. That was improper and the Court should—again—order Meta to respond to the RFA as written. If Meta can't admit the RFA using the Court's definition of "relevant copyright owner" for any particular owner, then it can deny the RFA as to that owner(s).

Meta also admits the request only as to "one or more Datasets" instead of answering it as written—i.e., about *all* of the datasets used to train Llama. *Id.* If Meta does not believe a particular dataset it used to train Llama included copyrighted books then it should say so and deny the RFA as to that dataset, but it must admit the request as to the rest, not "one or more," of them. Meta has the capability of answering directly; it already did so for Books3 in response to RFA 18. *See* Ex. A.

RFAs 16, 19, and 22 ask Meta to admit it used Books3, LibGen, and The Pile as datasets to train Llama. Rather than answer the requests as written, Meta rewrites them to ask whether Meta "used a portion of" Books3 and LibGen and "used some content" of The Pile. This is evasive. Meta's witnesses and documents confirm that Meta downloaded and used all three as datasets to train Llama. Meta should admit the requests as written. If Meta wants to deny as to some "portion" or "content," then it should identify that portion or content and deny the requests as to it.

RFA 26 asks about "copyright owners," not "Persons," and about "negotiat[ing] licensing of their copyrighted material," not "agreement[s] for access to and use of certain data that may include copyrighted material." Ex. A at 19-20. Meta should admit or deny the question as written.

RFA 35 asks Meta to admit it hasn't given Plaintiffs a list of works in the datasets it uses. Meta denies the RFA in full, stating it has provided a list from Books3, but then goes on to say it doesn't have lists from other datasets with some unexplained relevance and proportionality objections. Those objections are meritless. If Meta doesn't have any lists of works used in other datasets then obviously it hasn't provided Plaintiffs with them and should admit the RFAs.

RFA 39 asks Meta to admit it hasn't "deleted all copyrighted material" in its possession once used for training Llama. Rather than answer that question as posed, Meta rewrites it to say it hasn't deleted all "training data" and then tries to defend its decision not to do so by pointing to preservation obligations in a non-exhaustive list. The former is not what the request asks, and the latter tries to answer a question that wasn't asked. The Court should order Meta to answer the request as written or at minimum to state *all* the reasons it hasn't deleted the works, not just its preferred reason.

C. Meta Should Answer RFAs 38, 44, and 98.

Meta denies RFAs 38 ("Admit that You store copyrighted material for training Llama Models."), 44 ("Admit that if copyright holders or other content creators demanded that You not use their content to train Your LLM models, then You would not use their content to train Your LLM models."), and 98 ("Admit that you used books sourced from Books3 to train one or more of your large language models.") because it purportedly "does not understand" them. Ex. A. These requests are clear. But to the extent "store" in RFA 38 was unclear, Plaintiffs told Meta it meant "maintain a copy or copies of, however arranged in part or in whole." Meta still refuses to answer. And to the extent "sourced" in RFA 98 was unclear, Plaintiffs told Meta it could just delete that word. Meta still refuses to answer. The Court should order Meta to answer these RFAs.

D. Meta Should Withdraw Its "Except As Expressly Admitted" Qualifications.

For most of its RFA responses—specifically, RFAs 1-26, 28, 29, 31, 32, 34, 35, 39, 42, 43, 45-91, 94 and 96—Meta admits or denies the request and then adds "Except as expressly admitted, Meta denies the Request" or "lacks information sufficient to form a belief as to the truth of this request, and on that basis denies it." These rote qualifications are nonsensical—for example, Meta even adds it to its RFA 35 response even though it denies that RFA. The Court should order Meta to remove this wasteful boilerplate.

E. Meta Should Withdraw Its "Subject To And Without Waiving" Qualifications.

In dozens of responses, even for its clean admissions, Meta prefaces with the clause, "Subject to and without waiting the foregoing objections . . ." This boilerplate obscures Meta's responses and makes it unclear which objections limit which responses (if any). At the M&C, Meta's counsel asserted this vague language retains some unexplained significance—it does not.

II.   META'S POSITION

This is Plaintiffs' second challenge to Meta's RFA responses. Plaintiffs filed their first on November 26 (ECF No. 293), challenging Meta's responses to more than seventy RFAs (1, 3-13, 17-18, 20-21, 23-24, 34, 43, and 45-96). The Court denied the motion as to RFAs 10-13 and 92, 93, and 95. ECF No. 315 ("Order") at 4-5. As to the remainder, the Court ordered Meta to amend in certain specific respects, which Meta promptly did on December 13. Exs. A, B. Plaintiffs now seek to raise not only purported concerns with Meta's amended responses (which fully comply with the Order), but also new issues with the *original RFA responses* that they could have raised previously. Plaintiffs' attempted do-over is untimely and baseless. Meta has responded fully and appropriately to all RFAs. Plaintiffs are entitled to nothing further.

Plaintiffs' Motion Is Untimely and Procedurally Improper. In his order of October 4 (ECF No. 211), Judge Chhabria instructed: "The parties are urged to bring any significant discovery disputes to Judge Hixson's attention sooner rather than later." Meta served its responses to Plaintiffs' two sets of RFAs on November 8 and 18, and Plaintiffs moved to compel as to both on November 26. Plaintiffs' decision to wait a full, additional month to raise *different, serial challenges to the same RFA responses* violates Judge Chhabria's order. If Plaintiffs had a problem with Meta's original RFA responses, they should have briefed all of those issues promptly at the same time so that Meta could address any necessary amendments once. *See, e.g., In re W. States Wholesale Nat. Gas Antitrust Litig.*, 2016 WL 2593916, at *4 (D. Nev. May 5, 2016), *objections overruled*, 2016 WL 3965185 (D. Nev. July 22, 2016) ("The court will simply not tolerate an unending stream of supplemental papers on issues that have been or could have been fully briefed and argued."); *Safeco Ins. Co. of Am. v. Rawstrom*, 183 F.R.D. 668, 671 (C.D. Cal. 1998) ("Allowing a party to present objections … serially would be inefficient.… The result would be unacceptable delay in the completion of discovery, the unnecessary expense of serial meetings between counsel, and the unnecessary burden of serial motions to compel.").

Meta Properly Responded To RFAs 3-6, 17, 20, 23, 34, 43, 45–91, 94, and 96. **RFAs 3-6 and 34** asked Meta to admit that the data it used to train its AI models "included [unidentified] copyrighted books." **RFAs 17, 20, and 23** asked Meta to admit that certain datasets created by third parties "contain[] [unidentified] copyrighted works." **RFAs 45-89** asked Meta to admit that the "Books3 database contains" each of the original twelve Plaintiffs' at-issue books. **RFAs 90, 91, 94, and 96** asked a similar question for Plaintiff Farnsworth: "Admit that [title] by Christopher Farnsworth was included in a dataset used to train Your [LLMs]." Meta's responses admitted that the materials it used to train its models included text from copyrighted books or Plaintiffs' books.

Plaintiffs complain that Meta's responses did not address the books in their ***entirety***. The "reason Meta does so is obvious," but not for the reason Plaintiffs allege. As written, these RFAs would require an enormously burdensome word-by-word comparison not only for Plaintiffs' ***49*** books currently at issue, collectively spanning more than ***13,000*** pages, but for all other "copyrighted works" in numerous datasets. Responding to the RFAs as written would require, for each book or work, that Meta scan the entirety of the physical copy into digital form, identify and correct all scanning errors, account for artifacts such as page numbers, headers, formatting, etc., that are not part of the book content, and then perform a word-for-word comparison of that data against text in the training dataset used to train Llama. This process would involve a substantial amount of engineering work and require that Meta undertake an analysis more suitable to expert discovery—one that Plaintiffs are equally capable of doing.

Meta could have refused to answer on this basis alone, but it nevertheless *admitted* that text

from copyrighted books/Plaintiffs' books was present in the training datasets. Those responses were based on manually selecting samples of text across each of the Plaintiffs' 49 books at issue and running searches against the training data. That process involved a significant burden but nowhere near the word-for-word comparison that Plaintiffs demand. Further, to the extent these RFAs were seeking an admission that Meta trained any Llama model on the entirety of Plaintiffs' books, Meta has concluded based on a reasonable investigation that it did not do so. Thus, Meta properly admitted that it trained on some text from, but not the entirety of, each of Plaintiff's books. Fed. R. Civ. P. 36(a)(4) provides: "[W]hen good faith requires that a party qualify an answer or deny only a part of a matter, the answer must specify the part admitted and qualify or deny the rest." That is exactly what Meta did. No other response would be accurate or is required.

Plaintiffs also challenge Meta's response to **RFA 43**, which asked Meta to admit that its models "that were trained with copyrighted material had at least in part a commercial purpose." Meta properly admitted both that its training material included text from copyrighted works and that third parties can use Llama 2 and 3 "for commercial uses pursuant to certain terms and conditions." It's unclear why Plaintiffs believe this response to be inadequate. It is not.

Meta Properly Responded to RFA 7. **RFA 7** asked Meta to admit that it "did not obtain permission or consent from the relevant copyright owners to use all copyrighted books in the Datasets used to train Llama Models." In response, Meta noted that it has no way to confirm who the copyright owners are for each book (some of which may be out of copyright, subject to assignment of rights, etc.). Critically, however, Meta did ***not*** refuse to answer on this basis. Instead, it reasonably construed "relevant copyright owner" to mean the person "identified in the book as the copyright owner" and *admitted* that its datasets included text from "one or more copyrighted books for which it did not obtain permission or consent from the relevant copyright owner(s) (as construed above)." This admission was accurate, narrowly qualified, and responsive to the question asked—which was directed indiscriminately to whether Meta sought permission to use "*all*" of the many thousands of books in the training sets. Meta did not, and has said so. To the extent Plaintiffs are now asking Meta to provide an individualized admission or denial as to each of tens of thousands of (unidentified) books and authors—most of which are not at issue in this case—it is *Plaintiffs* who are improperly attempting to rewrite their RFA so as to be extraordinarily overbroad, unduly burdensome, and disproportionate to the needs of the case. Plaintiffs' further demands as to this RFA are untenable, unwarranted, and should be denied.[1]

Meta Properly Responded to RFAs 16, 19, 22, 26, 35, and 39. Plaintiffs could have— but failed to—move on any of these RFAs previously for good reason: Meta's responses are fully appropriate. **RFAs 16, 19, and 22** asked Meta to admit that it used certain third party datasets "to train one or more Llama Models." Meta admitted it used part of each dataset, and otherwise denied the RFAs—which is the only accurate response and fully appropriate under Rule 36(a)(4). An RFA is not an interrogatory and does not impose upon Meta any obligation to provide a narrative about what it did or did not use. In response to **RFA 26**, rather than denying outright, Meta admitted as much as it could, namely, that it had contacted persons to discuss agreements to access and use data that may include copyrighted material as training material. The rest of the RFA is inaccurate, and Meta denies it. **RFA 35** asked Meta to admit that it "has not provided Plaintiffs a list of works used in the Datasets used to train Llama Models." Meta denied this request and

---

[1] Plaintiffs do not challenge Meta's very similar response to RFA 18, admitting it did not "obtain authorization from all copyright owners (as construed above) of works included within … Books3 to use that dataset for training," making Plaintiffs' complaints about RFA 7 even more inscrutable.

explained why: it *has* provided such a list (and cited the specific production number). Finally, **RFA 39** asked Meta to admit that it has "not deleted all copyrighted material in [its] possession after it [wa]s used for training Llama Models." This is irrelevant and goes well beyond Plaintiffs' books, and Meta has in any event admitted that it did not delete all training material (consistent with its preservation obligations in this case). Nothing more is required, and certainly not the Interrogatory-like response Plaintiffs are demanding.

Meta's Denials to RFAs 38, 44, and 98 Should Stand. Meta explained more than a month ago that it did not understand what these RFAs were asking, denied on that basis, and offered a meet and confer, which Plaintiffs never sought until now. Plaintiffs' belated offer to rewrite the RFAs does not make them any clearer. Whether Meta "stores" or "maintains a copy of" copyrighted content for training (**RFA 38**) is still vague and overbroad since no such content is identified and this is not directed to Plaintiffs' books. Meta cannot admit or deny what it would do or not do in response to the incomplete hypothetical in **RFA 44**. And, even as modified by Plaintiffs, **RFA 98** remains vague and duplicative of many other RFAs (*see, e.g.,* 3, 4, 5, 6).

Meta Properly Qualified All Admissions and Denials. Meta's responses to **RFAs 1-26, 28, 29, 31, 32, 34, 35, 39, 42, 43, 45-91, 94 and 96** (many of which Plaintiffs failed to challenge previously) admit that part of each RFA is true and otherwise deny the rest. This is not "nonsensical"; it is *required* by Rule 36(a)(4). Likewise, Plaintiffs' complaint about standard prefatory language that they, themselves, used in their RFA responses (*see, e.g.*, Ex. B) is not well taken. All of Meta's RFA responses start by explaining what Meta objects to and why, including how it construes terms, and then answers "subject to" those qualifications. This language is required for accuracy and clarity on what Meta is admitting or denying, and Plaintiffs' request to remove it is unfounded (and should have been raised, if at all, in their earlier motion to compel).

III.   PLAINTIFFS' REPLY

Timeliness. Meta ignores Judge Chhabria's order that "the deadline to raise disputes regarding additional discovery"—which includes these RFAs—remains 7 days after the close of discovery[.]" Dkt. 253. Meta cites no case suggesting a party waives a challenge to deficient RFA responses merely because it could have raised earlier. Meta then complains Plaintiffs make "serial challenges," but the two cases Meta cites are entirely inapposite—*Western States* discusses serial briefing on a motion, and *Safeco* is about interposing new objections in amended responses.[2]

Books vs. Text. Meta is playing games. RFAs 3-6, 17, 20, 23, 34, 43, 45–91, 94, and 96 ask about copyrighted works. Meta cannot answer based only on "text." It also should not be permitted to backtrack on its representation that it *already* "compar[ed] the texts from the published, commercially-available versions of Plaintiffs' works"—not "some text"—against third-party datasets to determine whether the datasets included text from Plaintiffs' at-issue books[.]" Meta does not suggest *any* reason to believe that *any* text from *any* book, including Plaintiffs', was missing from what Meta downloaded from Shadow Datasets and used to train its Llama models.

RFA 7. Meta knows it didn't obtain permission from any copyright owners. It should stop pretending otherwise and answer the RFA using the definition already ordered in Dkt. 315.

---

[2] Due to page limits, Plaintiffs do not reply to Meta's responses to RFAs 16, 19, 22, 25, 26, 38, 39, 44, or 98. Suffice it to say, Plaintiffs believe simply reading the RFAs and Meta's responses makes clear that the responses are deficient. The Court should order Meta to comply with Rule 36(a)(4) by either admitting or denying the requests as written or stating with specificity what it admits and what it denies, and that's it.

By: /s/ *Bobby Ghajar*

Bobby A. Ghajar
Colette A. Ghazarian
COOLEY LLP
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400
Facsimile: (310) 883-6500
Email: bghajar@cooley.com
          cghazarian@cooley.com

Mark R. Weinstein
Elizabeth L. Stameshkin
COOLEY LLP
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000
Facsimile: (650) 849-7400
Email: mweinstein@cooley.com
          lstameshkin@cooley.com

Kathleen R. Hartnett
Judd D. Lauter
COOLEY LLP
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111
Telephone: (415) 693-2071
Facsimile: (415) 693-2222
Email: khartnett@cooley.com
          jlauter@cooley.com

Phillip Morton
COOLEY LLP
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004
Telephone: (202) 842-7800
Facsimile: (202) 842-7899
Email: pmorton@cooley.com

Angela L. Dunning
**CLEARY GOTTLIEB STEEN &
HAMILTON LLP**
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304

By: /s/ *Maxwell V. Pritt*

BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jischiller@bsfllp.com
jstein@bsfllp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com

Joshua I. Schiller (SBN 330653)
David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@bsfllp.com

Interim Lead Counsel for Plaintiffs

Telephone: (650) 815-4121
Facsimile: (650) 849-7400
Email: adunning@cgsh.com

Attorneys for Defendant Meta Platforms, Inc.

ATTESTATION PURSUANT TO CIVIL LOCAL RULE 5-1(h)

I hereby attest that I obtained concurrence in the filing of this document from each of the

other signatories. I declare under penalty of perjury that the foregoing is true and correct.

Dated: December 20, 2024   BOIES SCHILLER FLEXNER LLP

                                                                  /s/ Maxwell V. Pritt
                                                                  Maxwell V. Pritt
                                                                  Reed Forbush
                                                                  Jay Schuffenhauer

                                                                  Attorneys for Plaintiffs

8